

PCT

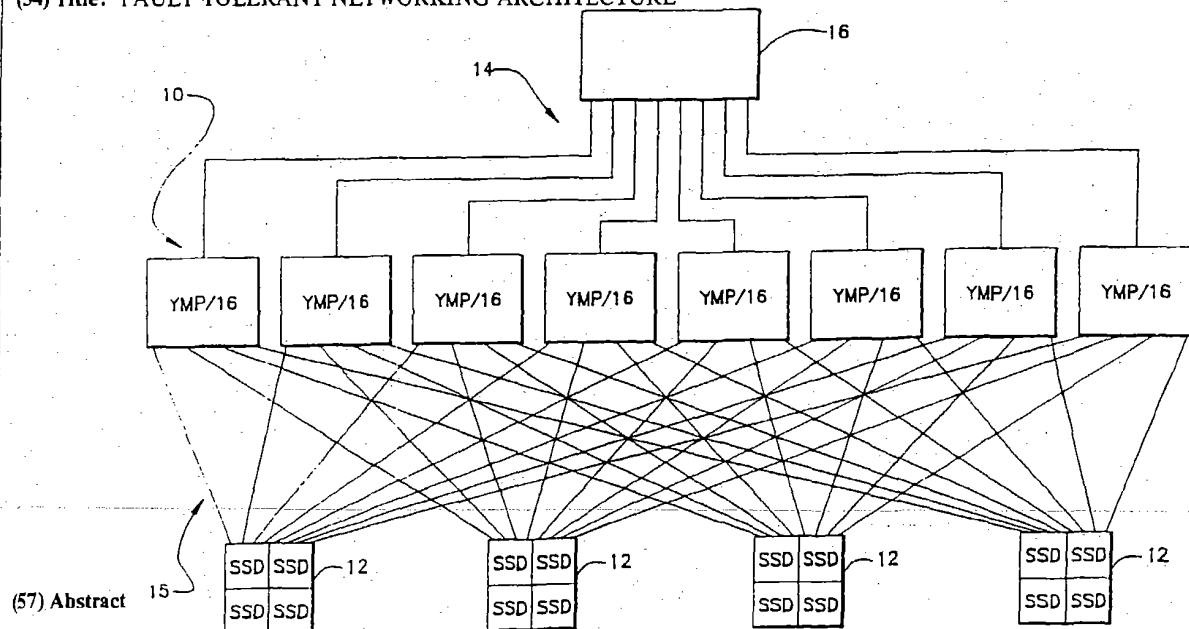
WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁵ : G06F 15/16, 11/18		A1	(11) International Publication Number: WO 92/04677
			(43) International Publication Date: 19 March 1992 (19.03.92)
(21) International Application Number: PCT/US91/06060 (22) International Filing Date: 23 August 1991 (23.08.91) (30) Priority data: 582,507 12 September 1990 (12.09.90) US (71) Applicant: CRAY RESEARCH, INC. [US/US]; 608 Second Avenue South, Minneapolis, MN 55402 (US). (72) Inventors: SUNDET, James, Wallace ; 17110 Regina Street, Chippewa Falls, WI 54729 (US). BROWN, Roger, Gene ; Route 3 Box 93A, Colfax, WI 54730 (US). (74) Agent: HAMRE, Curtis, B.; Merchant, Gould, Smith, Edell, Welter & Schmit, 3100 Norwest Center, 90 South Seventh Street, Minneapolis, MN 55402 (US).		(81) Designated States: AT (European patent), BE (European patent), CA, CH (European patent), DE (European patent), DK (European patent), ES (European patent), FR (European patent), GB (European patent), GR (European patent), IT (European patent), JP, KR, LU (European patent), NL (European patent), SE (European patent). Published <i>With international search report.</i>	

(54) Title: FAULT TOLERANT NETWORKING ARCHITECTURE



(57) Abstract

A fault tolerant network for a plurality of computers includes a system for controlling access to shared peripherals. Access to the shared peripherals is coordinated among the computers by means of communication through a semaphore box. Each computer connects to the semaphore box via a channel. The semaphore box is comprised of two major sections: a semaphore section and an I/O section. The semaphore section contains two sets of semaphores: a first set comprising reservation semaphores for the shared peripherals; and a second set comprising heartbeat semaphores for the sharing computers. The first set is used to reserve a particular peripheral for a particular computer and indicate the source of the reservation; the second set provides a "heartbeat" to prevent reservation semaphores from being set indefinitely in the event communication with a particular computer is lost.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	ES	Spain	MG	Madagascar
AU	Australia	FI	Finland	ML	Mali
BB	Barbados	FR	France	MN	Mongolia
BE	Belgium	GA	Gabon	MR	Mauritania
BF	Burkina Faso	GB	United Kingdom	MW	Malawi
BG	Bulgaria	GN	Guinea	NL	Netherlands
BJ	Benin	GR	Greece	NO	Norway
BR	Brazil	HU	Hungary	PL	Poland
CA	Canada	IT	Italy	RO	Romania
CF	Central African Republic	JP	Japan	SD	Sudan
CG	Congo	KP	Democratic People's Republic of Korea	SE	Sweden
CH	Switzerland	KR	Republic of Korea	SN	Senegal
CI	Côte d'Ivoire	LI	Liechtenstein	SU ⁺	Soviet Union
CM	Cameroon	LK	Sri Lanka	TD	Chad
CS	Czechoslovakia	LU	Luxembourg	TG	Togo
DE*	Germany	MC	Monaco	US	United States of America
DK	Denmark				

⁺ Any designation of "SU" has effect in the Russian Federation. It is not yet known whether any such designation has effect in other States of the former Soviet Union.

FAULT TOLERANT NETWORKING ARCHITECTURE

BACKGROUND OF THE INVENTION

5 1. Field Of The Invention.

This invention relates generally to computer networks, and in particular to a fault tolerant network having a semaphore box for controlling access to shared peripherals by a plurality of computers.

10

2. Description of Related Art.

Multiprocessor systems typically include some method of providing interprocessor communication. For example, interprocessor communication through a shared main memory is typically referred to as a "loosely coupled" computer system. Interprocessor communication through shared registers is typically referred to as a "tightly coupled" computer system. Prior patents of the Assignee of the present invention, Cray Research, Inc., disclose various forms of interprocessor communication.

20

One such prior patent is U.S. Pat. No. 4,636,942, issued January 13, 1987, to Chen et al., which patent is incorporated herein by reference. This patent discloses a computer vector multiprocessing control wherein a pair of processors are provided and each are connected to a central memory through a plurality of memory reference ports. Processors are further connected to a plurality of shared registers, including registers for holding scalar and address information, and registers for holding information to be used in coordinating the transfer of information through the shared registers.

25

30

Another prior patent is U.S. Pat. No. 4,661,900, issued April 28, 1987, to Chen et al., which patent is incorporated herein by reference. This patent discloses a flexible chaining method and apparatus wherein a pair of processors are connected to a central memory through a plurality of memory reference ports. The processors are further connected to a plurality of shared registers that may be directly addressed by either processor, and which

35

-2-

hold scalar and address information in registers for holding information to be used in coordinating the transfer of information through the shared registers.

Still another prior patent is U.S. Pat. No. 4,754,398, issued June 28, 1988, to Pribnow, which patent is incorporated herein by reference. This patent discloses an interprocessor communication system for a multiprocessor system that includes a plurality of clusters having a plurality of semaphore registers and information registers.

Whatever the merits of these prior patents for controlling interprocessor communication, they do not achieve the benefits of the present invention.

SUMMARY OF THE INVENTION

The present invention discloses a fault tolerant network for computers that includes a semaphore box for controlling access to shared peripherals. The semaphore box is comprised of two major sections: an I/O section and a semaphore section. The semaphore section contains two sets of semaphores: a first set comprising reservation semaphores for the shared peripherals and a second set comprising heartbeat semaphores for the sharing computers. The first set is used to reserve a particular peripheral for a particular computer; the second set provides a "heartbeat" to prevent reservation semaphores from being set indefinitely in the event communication with a particular computer is lost.

The heartbeat semaphores are arranged in an array. Each time a semaphore command is received, a row of heartbeat semaphores is set. When the command completes, a column of heartbeat semaphores is returned to the requesting computer and then cleared. If the heartbeat semaphore in a particular position of the column is set, then the associated computer has accessed the semaphore box sometime between the current access of the requesting computer and its prior access. Conversely, if the

-3-

particular heartbeat semaphore is cleared, then the associated computer has not accessed the semaphore box during the period. If the particular heartbeat semaphore for the associated computer remains cleared for some number
5 of consecutive accesses, the requesting computer should conclude that the associated computer has lost communication with the semaphore box.

BRIEF DESCRIPTION OF THE DRAWINGS

10 Referring now to the drawings, in which like reference numbers represent like elements throughout the several views:

Figure 1 is a block diagram of a system configuration according to the preferred embodiment;

15 Figure 2 is a block diagram of the semaphore box in the preferred embodiment;

Figure 3 illustrates a channel command word in the preferred embodiment; and

20 Figure 4 illustrates the format of the channel status word in the preferred embodiment.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT

In the following description of the preferred embodiment, reference is made to the accompanying drawings
25 which form a part hereof, and in which is shown by way of illustration a specific embodiment in which the invention may be practiced. It is understood that other embodiments may be used and structural changes may be made without departing from the scope of the present invention.

30

Glossary

In the following description, the term "semaphore" is often used. In the preferred embodiment, this term refers to a memory cell that is shared by a plurality of
35 processors to provide a form of communication between the

-4-

processors by indicating when significant events have taken place.

System Configuration

5 Figure 1 is a block diagram of a system configuration according to the preferred embodiment of the present invention. The present invention discloses a fault tolerant network for computers 10 that includes a semaphore box 16 for controlling access to shared peripherals 12.

10 The computers 10 may be YMP/16 computers of the type manufactured by Cray Research, Inc., the Assignee of the present invention. The peripherals 12 may be Solid-state Storage Devices (SSDs) of the type described in either U.S. Pat. No. 4,630,230, issued December 16, 1986, to Sundet
15 (which patent is incorporated herein by reference), or in U.S. Pat. No. 4,951,246, issued August 21, 1990, to Fromm et al. (which patent is incorporated herein by reference). Those skilled in the art, however, will recognize that other computers 10 and other peripherals 12 could
20 substituted therefor.

Each peripheral 12 connects to the computers 10 via a high speed channel 15. Each computer 10 connects to a semaphore box 16 via a low speed channel 14. Access to the shared peripherals 12 is coordinated among the computers 10
25 by means of communication through the semaphore box 16.

Semaphore Box

A block diagram of the semaphore box 16 is shown in Figure 2. The semaphore box 16 is comprised of two major
30 sections: a semaphore section and an I/O section.

The semaphore section contains two sets of semaphores: a first set comprising reservation semaphores for the shared peripherals 12; and a second set comprising heartbeat semaphores for the sharing computers 10. The
35 first set is used to reserve a particular peripheral 12 for a particular computer 10; the second set provides a

-5-

"heartbeat" to prevent reservation semaphores from being set indefinitely in the event communication with a particular computer 10 is lost. In the preferred embodiment, there are three identical copies, or groups 18, 20, and 22, of the two sets arranged in a triple module redundant configuration. The three groups are identified as: semaphore group A (18), semaphore group B (20), and semaphore group C (22).

The reservation semaphores can be individually set, cleared, or tested by commands received by the semaphore box 16 from the computers 10. In the preferred embodiment, each reservation semaphore is four bits long. One bit contains reservation information associated with one of the peripherals 12; the other three bits are a port identifier field containing a number of the input port 24A-24H which most recently set the reservation semaphore. The port identifier field thus identifies a particular computer 10 connected to the associated input port 24A-24H. The port identifier field is updated each time the reservation semaphore is set.

Table 1 is an illustration of an array of heartbeat semaphores, wherein each heartbeat semaphore is identified by a concatenated row-column value.

-6-

Table 1

		COLUMN							
		0	1	2	3	4	5	6	7
5	ROW 0	00	01	02	03	04	05	06	07
	ROW 1	10	11	12	13	14	15	16	17
10	ROW 2	20	21	22	23	24	25	26	27
	ROW 3	30	31	32	33	34	35	36	37
	ROW 4	40	41	42	43	44	45	46	47
15	ROW 5	50	51	52	53	54	55	56	57
	ROW 6	60	61	62	63	64	65	66	67
20	ROW 7	70	71	72	73	74	75	76	77

The array as shown in Table 1 measures eight by eight, but those skilled in the art will recognize that any size array may be used. The array of heartbeat semaphores as shown in Table 1 is used to prevent the reservation semaphores from remaining in a "set" condition after communication with a computer 10 has been lost. A reservation semaphore that remains "set" prevents access to the associated peripheral 12 by the other computers 10.

Each time a semaphore command is received by the semaphore box 16, a row of heartbeat semaphores is set. Preferably, the row number is the same as the number of the input port 24A-24H which received the command (i.e. row N is set by any command from Port N), although other means of associating rows with computers 10 could be used. When the command completes, a column of heartbeat semaphores is

-7-

returned to the requesting computer 10 by the output ports 26A-26H and then cleared to all zeros. The column becomes part of the channel status word. Like the row number, the column number is preferably the same as the number of the input port 24A-24H which received the command (i.e. column N is returned in status word for Port N), although other means of associating columns with computers 10 could be used.

The heartbeat semaphores provide a means of determining whether the computers 10 connected to the semaphore box 16 are still active. This is best illustrated by the following example. Assume computer A wants to determine the running status of computer B. Each time computer B sends a command to the semaphore box 16, all heartbeat semaphores in row B are set. Each time computer A sends a command to the semaphore box 16, the heartbeat semaphores in column A are returned in the channel status word and then cleared. Since the rows and columns intersect, the Bth heartbeat semaphore in column A indicates whether computer B has accessed the semaphore box 16 in the period since the previous computer A access. If the heartbeat semaphore is set, computer B has accessed the semaphore box 16 sometime between the current access of computer A and the prior access of computer A. Conversely, if the heartbeat semaphore is clear, computer B has not accessed the semaphore box 16 during the period. If the heartbeat semaphore for computer B is clear for some number

-8-

of consecutive accesses or some predetermined interval, then computer A concludes that computer B has lost communication with the semaphore box 16 and takes appropriate action. Those skilled in the art will
5 recognize that the number of consecutive accesses or the predetermined interval is programmable. Thus, any computer 10 connected to the semaphore box 16 must periodically access the semaphore box 16 to keep its "heartbeat" alive.

Referring again to Figure 2, the I/O section of the
10 semaphore box 16 consists of input ports, voting circuits, and output ports. Each attached computer 10 communicates with the semaphore box 16 across a low speed channel 14 which attaches to the semaphore box 16 at an input port 24A-24H and an output port 26A-26H. Preferably, the
15 computer 10 is attached to output ports 26A-26H identically numbered or otherwise associated with input ports 24A-24H. Each input port 24A-24H and output port 26A-26H of the semaphore box 16 is logically independent from the others. Operations on the semaphore box 16 are accomplished by
20 transmitting commands to the input ports 24A-24H. The input ports 24A-24H then transmit the commands to all three of the semaphore groups 18, 20, and 22. (Note that in Figure 2 only a portion of the connections between the
input ports 24A-24H and the semaphore groups 18, 20, and 22
25 are illustrated.) The semaphore groups 18, 20, and 22 transmit the results of the operation to voting circuits 28A-28H. The three copies of the results are inspected by

-9-

voting circuits 28A-28H. The voting circuits 28A-28H detect an error if there is a difference in the execution results. If a single semaphore group 18, 20, or 22 is in error, its results are ignored by the voting circuits 28A-28H. If two semaphore groups 18, 20, or 22 fail, all comparisons will fail and the results of group 20 are used (although there is no assurance that the results of group 20 are correct). The output ports 26A-26H then transmit the results to the computer 10.

10 A port conflict occurs when more than one command is received by the semaphore section from the I/O section in the same clock period. Conflicts are resolved on a priority basis in the preferred embodiment, although other methods of resolving conflicts could be substituted
15 therefor. In the preferred embodiment, lower numbered input ports 24A-24H have priority over higher numbered input ports 24A-24H. Thus, requests from higher numbered input ports 24A-24H are held until the semaphore response for the lower numbered input ports 24A-24H is returned to
20 the I/O section. The requests are not queued, although alternative embodiments could easily implement such a scheme. Instead, in the preferred embodiment, if a lower numbered request is received while a higher number request is being held, the lower numbered request will again be
25 honored first.

-10-

Operation

Figure 3 illustrates a channel command word 32 in the preferred embodiment. The codes used in Figure 3 are defined as shown in Table 2.

5

Table 2

	CODE	MEANING
10	Fn	function code bit 2 ⁿ
	Sn	semaphore select code bit 2 ⁿ
15	xx	unused bit

The channel command word 32 consists of a single 64-bit word. It is transmitted over the low speed channel 14 as four 16-bit parcels followed by a channel disconnect pulse.

20 The parcels of the channel command word 32 are identified as: parcel 0 (32A), parcel 1 (32B), parcel 2 (32C), and parcel 3 (32D). After an input port 24A-24H receives the channel command word 32, it performs error checks to detect possible parity errors, command compare errors, or channel

25 protocol errors. Detection of a channel error aborts the operation.

Each parcel of the channel command word 32 contains the same information, but the individual bit assignments differ in each parcel. Rearranging the bit assignments

30 prevents a defective wire, connector pin, or associated circuits from causing a catastrophic error.

-11-

Eight bits of each parcel are used to hold a command. The remaining eight bits are unused, except for parity, and may contain any value. Of the eight command bits, two bits contain a semaphore function code and six bits contain a semaphore select code. The four possible function codes are described as shown in Table 3.

Table 3

CODE			OPERATION
	F1	F0	
0	0	0	Test semaphore.
0	1	0	Set semaphore unconditionally.
1	0	1	Clear semaphore unconditionally.
1	1	1	Test and set semaphore if clear. No operation if semaphore is set.

The semaphore select code determines which of the reservation semaphores are affected by the semaphore function code.

Each parcel is sent to a different semaphore group for execution: parcel 0 (32A) is sent to semaphore group 18; parcel 1 (32B) is sent to semaphore group 20; parcel 2 (32C) is sent to semaphore group 22; parcel 3 (32D) is discarded. Upon completion, the appropriate output port 26A-26H receives results from each of the three semaphore groups 18, 20, and 22. Both reservation semaphore and heartbeat semaphore information is contained in the

-12-

results. The three copies of the results are inspected by voting circuits 28A-28H and an appropriate channel status word is transmitted by the output ports 26A-26H to the originating computer 10. If the command received from the
5 input ports 24A-24H had been in error, the voting circuits 28A-28H detect the error and prevent corruption of any data on the shared peripherals 12.

Figure 4 illustrates the format of the channel status word 36, which consists of two 16-bit parcels followed by
10 a channel disconnect pulse. The parcels of the channel status word 36 are identified as: parcel 0 (36A), and parcel 1 (36B). The codes of the channel status word 36 in Figure 4 are defined as shown in Table 4, and the format is further described below.

-13-

Table 4

	CODE	MEANING
5	T0	test condition.
	S0	semaphore state after function.
	I0-2	port number of last reservation.
10	H0-7	heartbeat column bit 2 ⁿ .
	A0	any error.
15	C2	group C compare error.
	C1	group B compare error.
	C0	group A compare error.
20	P5	command compare error.
	P4	channel error.
25	p0-3	channel parity error, group n.
	P0-3	channel parity error, group n.
30	TM	test mode status bit.

Reservation Semaphore Status (T-, S-, I-)

The uppermost bit (T0) of the channel status word 36 is used only for a test and set operation. It reflects the state of the selected reservation semaphore at the time the function code is received by the semaphore section. Bit T0 is set to the value 1 if the reservation semaphore is initially set. Otherwise, a value of 0 is returned in this position if the reservation semaphore is initially clear, and the set operation has been performed.

Bit S0 reflects the condition of the selected reservation semaphore at the start of the current command.

-14-

Bits I0-I2 indicate the number of the input port 24A-24H (having a value of 0-7) which most recently changed the state if the selected reservation semaphore.

5 Heartbeat Status (H-)

The heartbeat status byte provides each computer 10 with information about the availability of the other computers 10 connected to the semaphore box 16. The bit position (H0-H7) of the heartbeat status corresponds to the 10 input port 24A-24H (0-7) which controls the setting of the particular heartbeat semaphore.

Port Status (A-, C-, P-, TM)

Bit A0 is 1 if an error is detected by the input ports 15 24A-24H or output ports 26A-26H. Bit A0 is a summation of bits C2-C0, P5, P4, p3-p0, and P3-P0 further described below.

Bits C0-C2 (semaphore compare error) are set if there is an error in the execution results. The comparison is 20 performed on 13 bits: the selected 4-bit semaphore, a semaphore test flag, and the 8-bit heartbeat semaphore column. A semaphore compare error could be caused by any of several conditions; a hardware malfunction; a command error; residue from a previous command error; or a 25 semaphore group 18, 20, or 22 whose contents have not been fully restored following a power loss or maintenance action. If a single semaphore group 18, 20, or 22 is in

-15-

error, its results are ignored by the voting circuits 28A-28H. If two semaphore groups 18, 20, or 22 fail, all comparisons will fail and the results of group 20 are used (although, there is no assurance that the results of group 5 20 are correct).

Bit C0 is set if information from semaphore group 18 fails to compare with either group 20 or 22. Similarly, bit C1 is set if information from semaphore group 20 fails to compare with either group 18 or 22. Bit C2 is set if 10 information from semaphore group 22 fails to compare with either group 18 or 20.

Bits P5-P0 indicate an error was detected during receipt of the command by the input ports 24A-24H. If P5-P0 are set, the command is aborted and parcel 0 (36A) of 15 the channel status word 36 is invalid.

Bit P5 is set if the function codes do not match in all four parcels of the channel command word 32. This could be the result of a programming error or a data error in the channel command word 32. The requested operation is 20 aborted and the command is not sent to the semaphore groups 18, 20, and 22. In the channel status word 36, the P5 bit (command compare error) is then set and parcel 0 (36A) of the channel status word 36 is invalid.

Bit P4 indicates that a channel protocol error was 25 detected by the input ports 24A-24H or output ports 26A-26H. Normal protocol is defined as four ready pulses each with its accompanying parcel of data followed by a

-16-

disconnect pulse. A resume pulse must be returned before a subsequent ready pulse is received. If more than four ready pulses are received before a disconnect pulse, then a channel protocol error has occurred. If a channel
5 protocol error is detected, then all command data is assumed to have been corrupted and is ignored. The state of the heartbeat semaphores is not changed and information in parcel 0 (36A) of the channel status word 36 is invalid. Bit P4 in the channel status word 36 is set. However, no
10 channel status word 36 is transmitted until a disconnect pulse is received. This ensures that any extra command parcels are flushed.

Additionally, the appropriate parity bits p3-p0 and P3-P0 are set in the event of a data error. Four parity
15 bits accompany each parcel of the channel command word 32. Bits P3-P0 perform a parity check using true logic levels, while bits p3-p0 perform a parity check using false logic levels. Thus, independent redundant checks are performed. Additionally, the P5 bit (command error) is set if a data
20 bit rather than a parity bit caused the parity error.

Bit TM is set to 1 when the output ports 26A-26H have been placed in test mode.

Other Features

25 The semaphore box 16 was designed for fault tolerance. For example, the semaphore box 16 may remain online and operational during repairs without impacting the integrity

-17-

of the semaphores. Further, even if two of the three semaphore groups 18, 20, and 22 fail, the semaphore box 16 continues operating using the remaining group. As mentioned above, if no group compares correctly with any other, group 20 is used.

All logic modules are supplied with a common clock from a master clock module and operate synchronously with one another. The master clock module is duplicated on a second module for backup purposes, but only one clock module may be powered on at any given time. Each logic module is provided with two clock inputs, one for each clock module. Selection of the active master clock module is by means of a manual switch which controls power to the clock module and provides the logical clock enable signals.

During the process of switching from one clock module to another, however, the system clock is not valid and the system requires re-initialization.

The semaphore box 16 resides in a standalone cabinet with its own cooling and internal power. Cooling is accomplished with forced room air. Multiple fans provide redundancy so that if a single fan malfunctions the equipment still remains operational.

Power is provided by four sets of identical power supplies to enhance fault tolerance. Each set of supplies is sized so that it is able to supply all the necessary power needs independently of the other. Power load shifting from the loss of one supply is automatic.

-18-

Each input port 24A-24H, output port 26A-26H, semaphore group 18, 20, and 22, and the master clock are tied to a common power bus, so that they may be individually disconnected from the power bus. Thus, no more than one module is affected by the loss of a single power supply breaker. Further, if power is removed from a single module, all other modules remain operational. Except for the master clock module, the process of applying or removing power to a single module does not affect the operation of the other modules. In addition, if a single module is removed from the system, all other modules remain operational. Thus, the process of inserting or removing modules does not affect the operation of the other modules.

15 Summary

In summary, a fault tolerant network has been described which includes a semaphore box 16 for controlling access to shared peripherals 12. The semaphore box 16 is comprised of two major sections: an I/O section and a semaphore section. The semaphore section contains reservation semaphores and heartbeat semaphores. The reservation semaphores are used to reserve a particular peripheral 12 for a particular computer 10; the heartbeat semaphores prevent reservation semaphores from being set indefinitely in the event communication with a particular computer 10 is lost.

-19-

The heartbeat semaphores are arranged in an array, wherein each time a command is received, a row of heartbeat semaphores is set. Further, when the command completes, a column of heartbeat semaphores is returned to the requesting computer 10 and then cleared. Thus, if the heartbeat semaphore in a particular position of the column is set, then the associated computer 10 has accessed the semaphore box 16 sometime between the current access of the requesting computer 10 and its prior access. Conversely, if the particular heartbeat semaphore is cleared, then the associated computer 10 has not accessed the semaphore box 16 during the period. The requesting computer 10 should conclude that the associated computer 10 has lost communication with the semaphore box 16 if the particular heartbeat semaphore for the associated computer 10 remains cleared for some number of consecutive accesses or some predetermined interval.

Conclusion

The foregoing description of the preferred embodiment of the present invention has been presented for the purposes of illustration and description. It is not intended to be exhaustive or to limit the invention to the precise form disclosed. Many modifications and variations are possible in light of the above teaching. It is intended that the scope of the invention be limited not by this detailed description, but rather by the claims appended hereto.

-20-

WHAT IS CLAIMED IS:

1. An apparatus for controlling access to at least one shared peripheral by a plurality of computers, comprising:
5 at least one reservation semaphore for reserving the peripheral for an accessing computer; and
at least one heartbeat semaphore operatively connected to the reservation semaphore for indicating whether the accessing computer has communicated with the apparatus
10 during some time period, thereby preventing the reservation semaphore from being reserved indefinitely when communications are lost with the accessing computer occurs.
2. The apparatus of claim 1, further comprising means for
15 arranging the reservation semaphores and the heartbeat semaphores in a redundant configuration to enhance fault tolerance.
3. The apparatus of claim 2, further comprising voting
20 circuit means for inspecting an outcome of an operation performed against the redundant configuration and detecting whether an error has occurred.
4. The apparatus of claim 1, wherein the reservation
25 semaphore comprises:
first means for storing a first indicator of whether the peripheral is reserved; and

-21-

second means for storing a second indicator of which computer reserved the peripheral.

5. The apparatus of claim 4, wherein the second means further comprises means for updating the second indicator each time the reservation semaphore changes state.

6. The apparatus of claim 1, further comprising:

means for transmitting a first plurality of the heartbeat semaphores to the accessing computer; and

means for examining the first plurality of the heartbeat semaphores to determine whether all of the computers connected to the apparatus are still active.

7. The apparatus of claim 1, further comprising:

means for arranging the heartbeat semaphores in a two-dimensional array comprising a plurality of intersecting rows and columns;

means for setting the heartbeat semaphores in a row of the array associated with the accessing computer whenever the accessing computer communicates with the apparatus;

means for transmitting to the accessing computer and then clearing the heartbeat semaphores in a column of the array associated with the accessing computer whenever the accessing computer communicates with the apparatus; and

means for determining when a disconnected computer has lost communications with the apparatus by examining the

-22-

heartbeat semaphore at an intersection of the column associated with the accessing computer and a row associated with the disconnected computer, wherein the heartbeat semaphore at the intersection remains cleared for a
5 predetermined period.

8. A method for controlling access to a shared peripheral by a plurality of computers, comprising:

setting a reservation semaphore to reserve the
10 peripheral for an accessing computer;

clearing the reservation semaphore to indicate the peripheral is available;

manipulating a heartbeat semaphore to indicate whether communication with the accessing computer has occurred,
15 thereby preventing the reservation semaphore from being reserved indefinitely when communications have been lost with the accessing computer.

9. The method of claim 8, further comprising arranging
20 the reservation semaphores and the heartbeat semaphores in a redundant configuration.

10. The method of claim 9, further comprising inspecting
an outcome of an operation performed against the redundant
25 configuration to detect whether an error has occurred.

-23-

11. The method of claim 8, wherein the setting step comprises:

storing a first indicator in the reservation semaphore of whether the peripheral is reserved; and

5 storing a second indicator in the reservation semaphore of whether the accessing computer has reserved the peripheral.

12. The method of claim 11, wherein the second indicator
10 storing step further comprises updating the second indicator each time the reservation semaphore changes state.

13. The method of claim 8, wherein the clearing step
15 comprises clearing the first indicator in the reservation semaphore when the peripheral is not reserved.

14. The method of claim 8, further comprising:
transmitting a first plurality of the heartbeat
20 semaphores to the accessing computer; and
examining the first plurality of the heartbeat semaphores to determine whether the computers connected to the apparatus are still active.

25 15. The method of claim 8, further comprising:
arranging the heartbeat semaphores in a

-24-

two-dimensional array comprising a plurality of intersecting rows and columns;

setting the heartbeat semaphores in a row of the array associated with the accessing computer;

5 transmitting and then clearing the heartbeat semaphores in a column of the array associated with the accessing computer whenever the accessing computer communicates with the apparatus; and

10 determining when a disconnected computer has lost communication with the apparatus by examining the heartbeat semaphore at an intersection of the column associated with the accessing computer and a row associated with the disconnected computer.

15

-24-

two-dimensional array comprising a plurality of intersecting rows and columns;

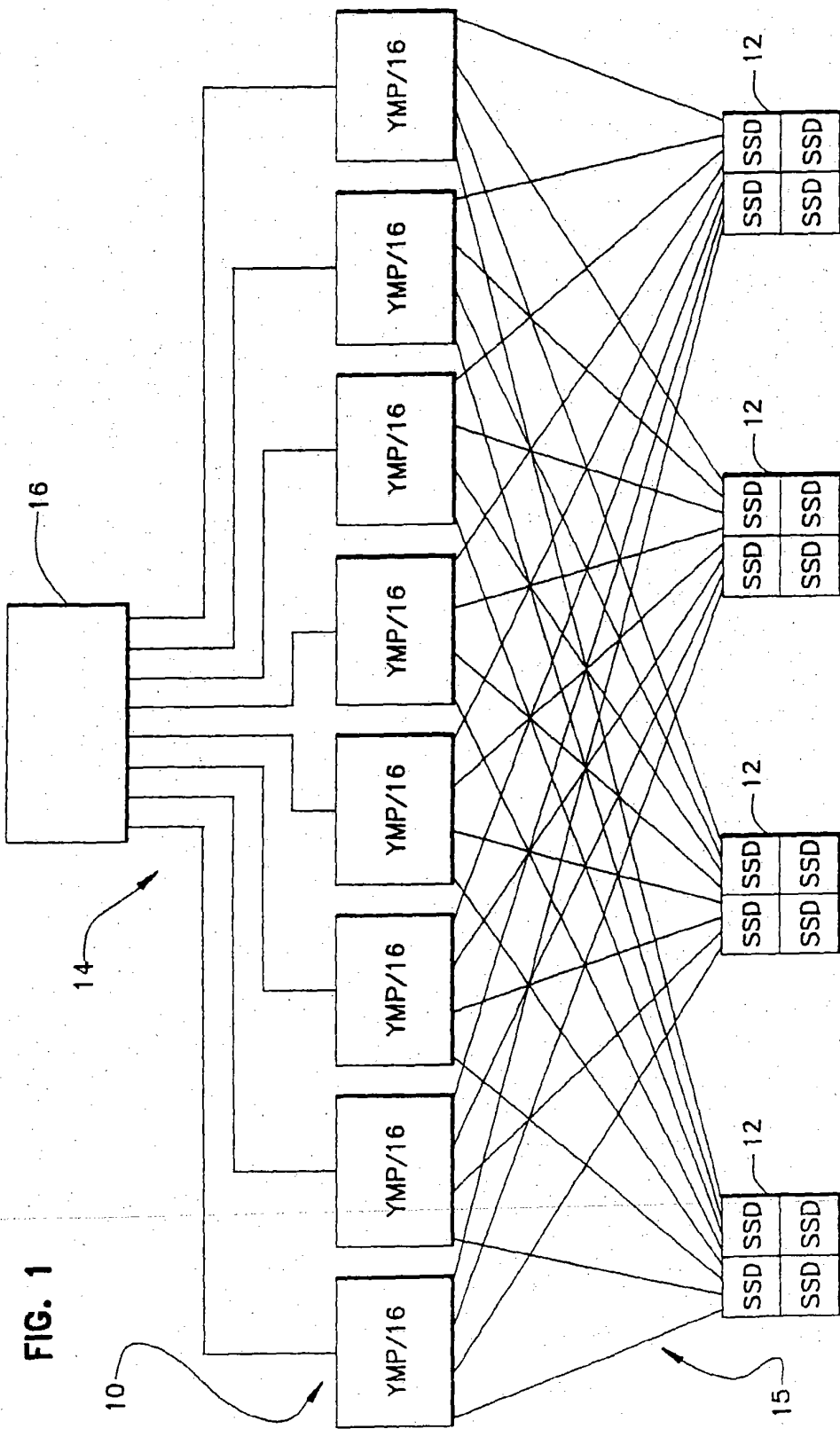
setting the heartbeat semaphores in a row of the array associated with the accessing computer;

5 transmitting and then clearing the heartbeat semaphores in a column of the array associated with the accessing computer whenever the accessing computer communicates with the apparatus; and

10 determining when a disconnected computer has lost communication with the apparatus by examining the heartbeat semaphore at an intersection of the column associated with the accessing computer and a row associated with the disconnected computer.

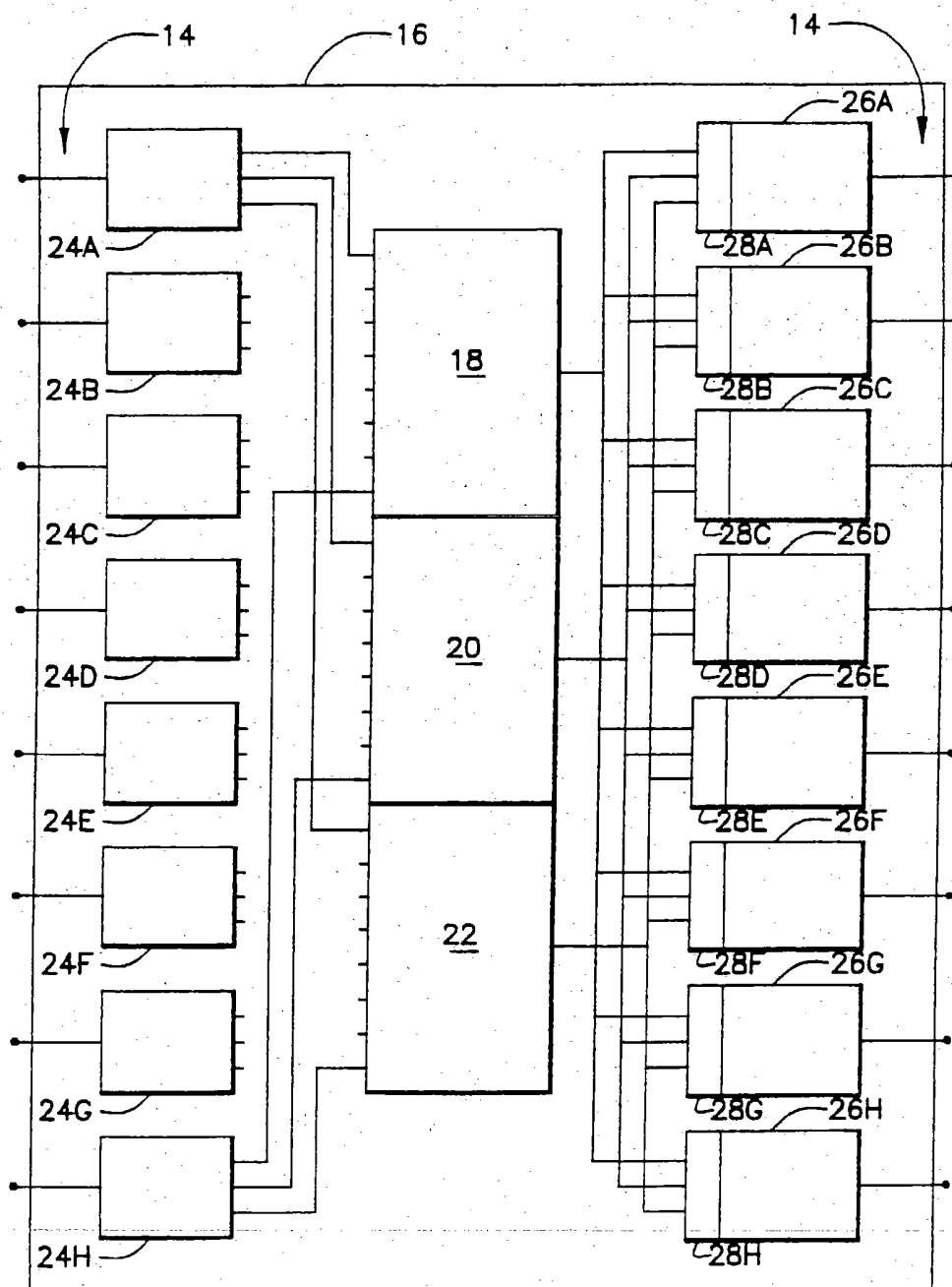
15

- 1 / 4 -

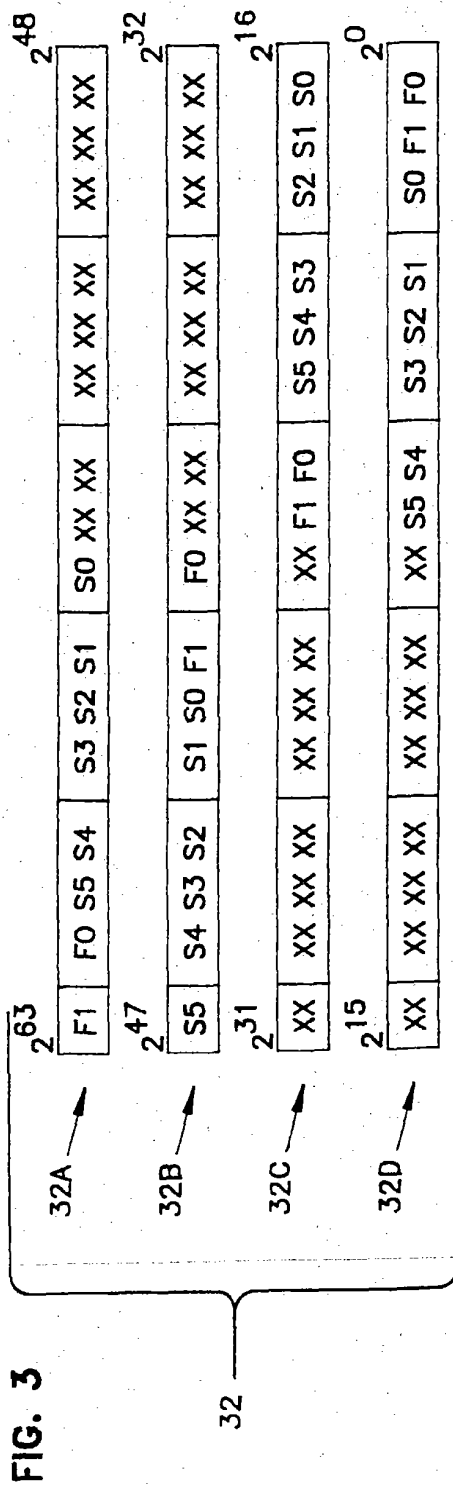


2/4

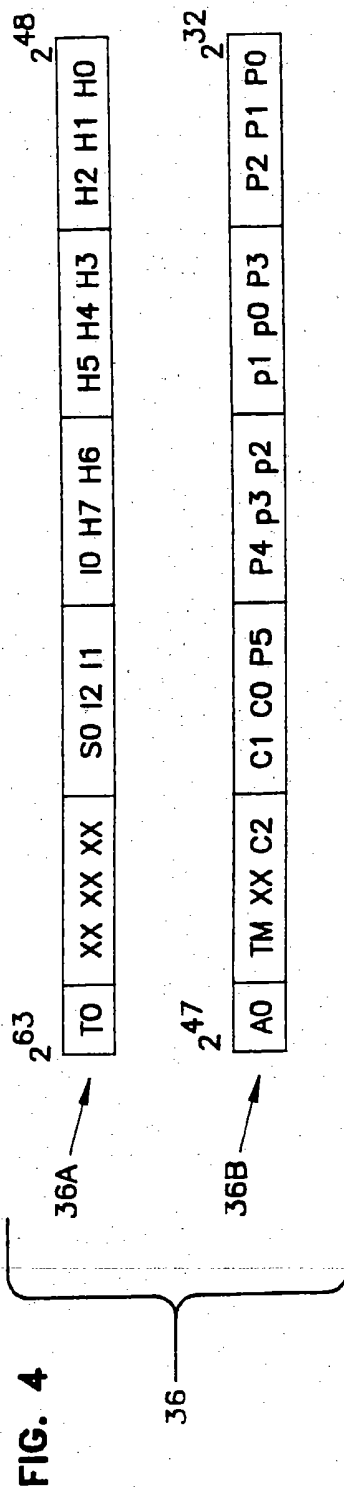
FIG. 2



3/4



4/4

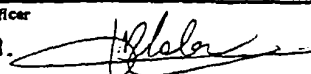


INTERNATIONAL SEARCH REPORT

International Application No.

PCT/US 91/06060

I. CLASSIFICATION OF SUBJECT MATTER (if several classification symbols apply, indicate all) ⁶		
According to International Patent Classification (IPC) or to both National Classification and IPC		
Int.Cl. 5 G06F15/16; G06F11/18		
II. FIELDS SEARCHED		
Minimum Documentation Searched ⁷		
Classification System	Classification Symbols	
Int.Cl. 5	G06F	
Documentation Searched other than Minimum Documentation to the Extent that such Documents are Included in the Fields Searched ⁸		
III. DOCUMENTS CONSIDERED TO BE RELEVANT⁹		
Category ¹⁰	Citation of Document, ¹¹ with indication, where appropriate, of the relevant passages ¹²	Relevant to Claim No. ¹³
A	US,A,4 754 398 (R.D. PRIBNOW) 28 June 1988 cited in the application see the whole document ---	1-15
A	EP,A,0 230 029 (A.T.T.) 29 July 1987 see column 2, line 5 - column 3, line 50 see abstract; claims; figures 6-9 ---	1,8
A	DE,A,2 057 030 (VEB KOMBINAT ROBOTON) 12 August 1971 see the whole document ---	1-4,8

	---/---	
<p>¹⁰ Special categories of cited documents:</p> <p>"A" document defining the general state of the art which is not considered to be of particular relevance</p> <p>"E" earlier document but published on or after the international filing date</p> <p>"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)</p> <p>"O" document referring to an oral disclosure, use, exhibition or other means</p> <p>"P" document published prior to the international filing date but later than the priority date claimed</p> <p>"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention</p> <p>"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step</p> <p>"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.</p> <p>"A" document member of the same patent family</p>		
IV. CERTIFICATION		
Date of the Actual Completion of the International Search	Date of Mailing of this International Search Report	
2 20 DECEMBER 1991	30, 12, 91	
International Searching Authority	Signature of Authorized Officer	
EUROPEAN PATENT OFFICE	SOLER J.M.B. 	

III. DOCUMENTS CONSIDERED TO BE RELEVANT (CONTINUED FROM THE SECOND SHEET)		
Category *	Citation of Document, with indication, where appropriate, of the relevant passages	Relevant to Claim No.
A	I & CS - INDUSTRIAL AND PROCESS CONTROL MAGAZINE. vol. 60, no. 10, October 1987, RADNOR, PENNSYLVANIA US pages 73 - 76; J.A. HUMPHRY: 'Applying fault tolerant system architectures' see the whole document ---	1-3
A	PATENT ABSTRACTS OF JAPAN vol. 8, no. 237 (P-310)(1674) 30 October 1984 & JP,A,59 113 600 (NIPPON DENKI KK) 30 June 1984 see the whole document ---	1,8
A	EP,A,0 035 778 (THE BOING COMPANY) 16 September 1981 see page 1, line 1 - page 10, line 10 see claims; figures 1-7 ---	7

**ANNEX TO THE INTERNATIONAL SEARCH REPORT
ON INTERNATIONAL PATENT APPLICATION NO. US 9106060
SA 51414**

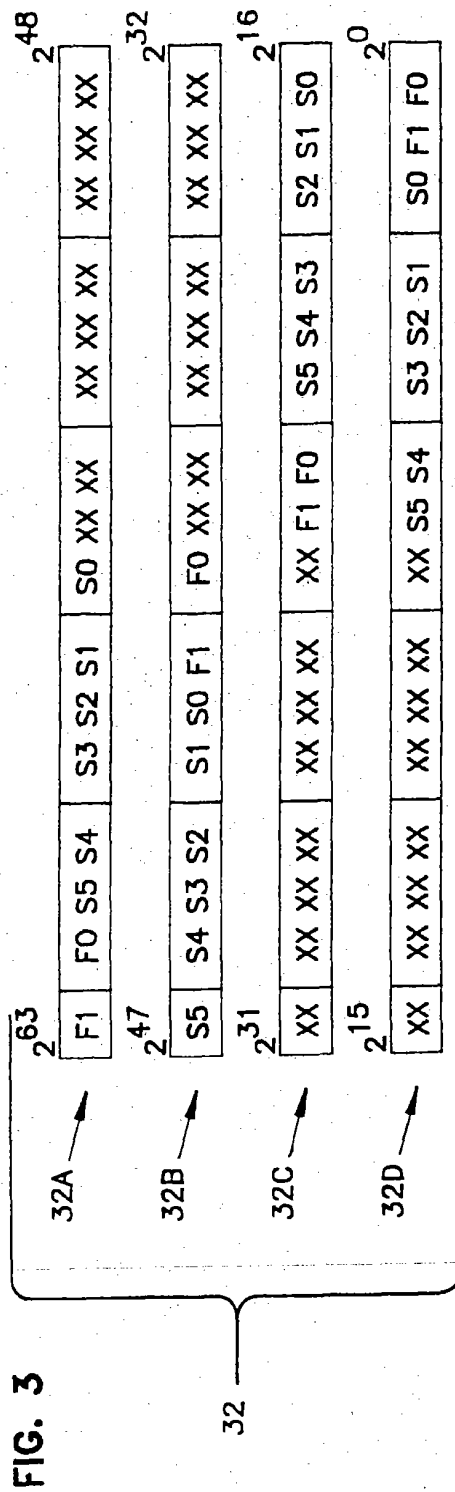
This annex lists the patent family members relating to the patent documents cited in the above-mentioned international search report.
The members are as contained in the European Patent Office EDP file on
The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information. 20/12/91

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US-A-4754398	28-06-88	CA-A- 1256582	27-06-89
EP-A-0230029	29-07-87	US-A- 4710926	01-12-87
		CA-A- 1267226	27-03-90
		JP-A- 62177634	04-08-87
DE-A-2057030	12-08-71	None	
EP-A-0035778	16-09-81	US-A- 4379326	05-04-83
		CA-A- 1156766	08-11-83
		WO-A- 8102645	17-09-81

EPO FORM P007

For more details about this annex : see Official Journal of the European Patent Office, No. 12/82

3/4



2/4

FIG. 2

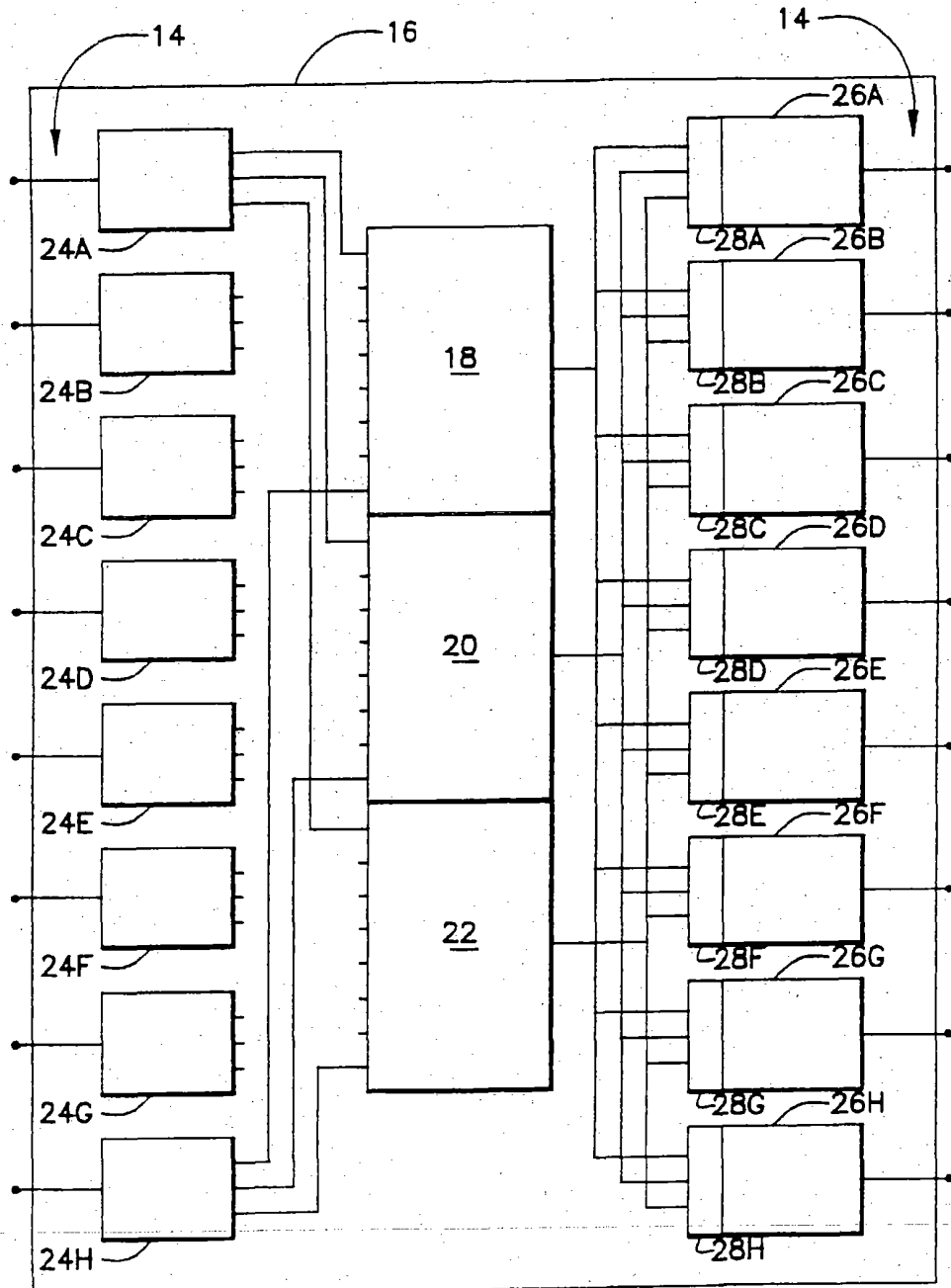
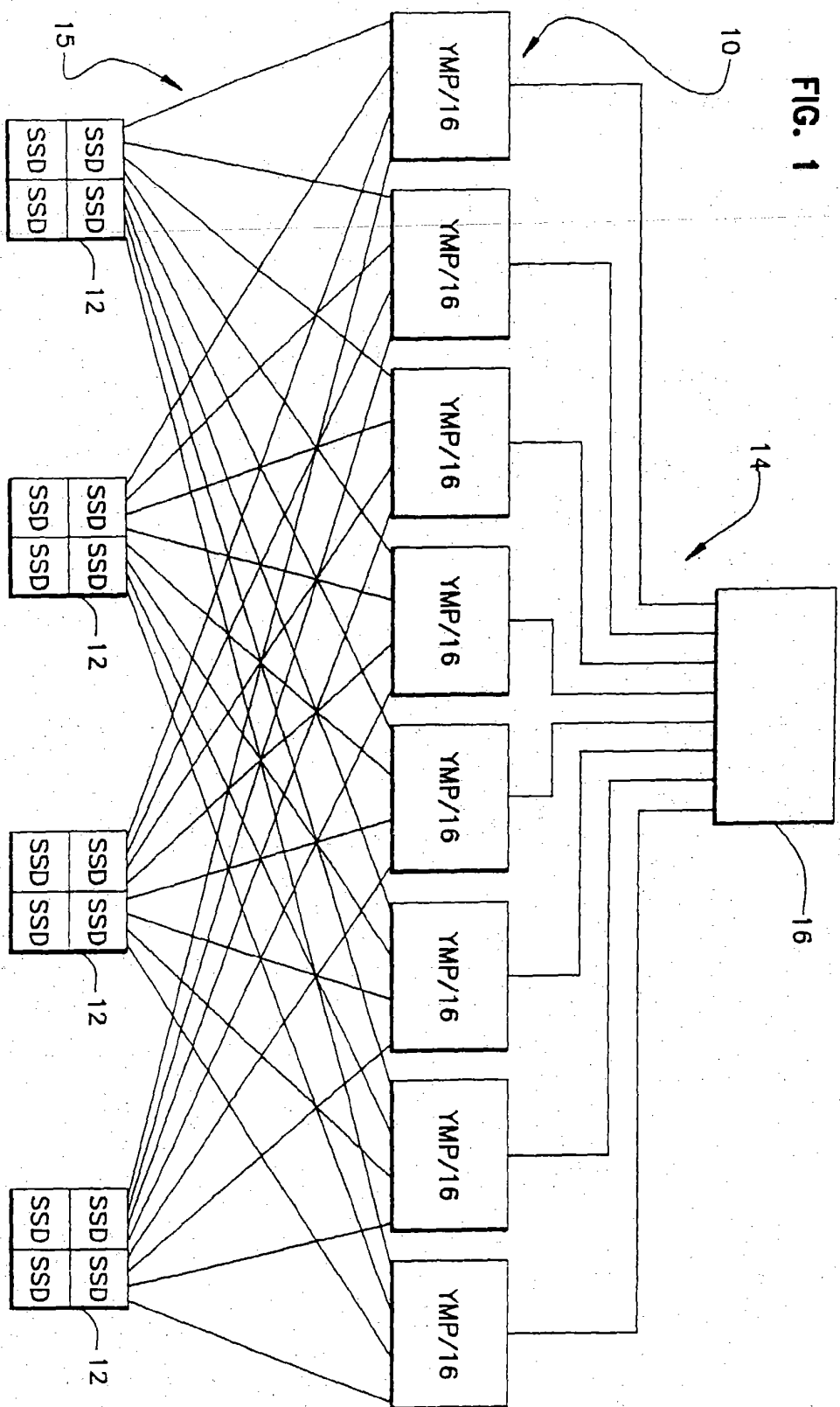


FIG. 1



- 1 / 4 -